DOCUMENT RESUME

ED 182 140 SE 029 850

AUTHOR Sadowski, Barbara R.
TITLE The Application of an Extreme-Types Model to a Domain-Referenced Diagnostic Arithmetic Test.
PUB DATE [78]
NOTE 17p.; Contains occasional broken type

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Diagnostic Tests; Educational Diagnosis; *Educational Research; Elementary Education; *Elementary School Mathematics; Group Tests; Test Reliability; Test Reviews; *Test Validity

ABSTRACT

The purpose of this research was to investigate the diagnostic validity of the Maryland Diagnostic Arithmetic Test (MDAT) as a group-administered test. The MDAT is a domain-referenced diagnostic test consisting of several sub-tests covering the content of elementary school arithmetic from sets to non-negative rational numbers. Forty-five non-overlapping domains were investigated using the extreme types model to determine homogeneous domains. Twelve other domains which were determined a priori were also investigated. The results of the group-administered MDAT provide some evidence for the validity of the MDAT as a diagnostic test. (Author/MK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Application of An Extreme-Types Model to a Domain-Referenced Diagnostic Arithmetic Test

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Barbara R. Sadowski

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Barbara R. Sadowski
University of Houston

ED182140

The purpose of this research was to investigate the diagnostic validity of the Maryland Diagnostic Arithmetic Test as a group-administered test. The Maryland Diagnostic Arithmetic Test (MDAT) is a domain-referenced diagnostic test consisting of several sub-tests covering the content of elementary school arithmetic from sets to non-negative rational numbers. The sub-tests which were researched as part of this study (Sadowski, 1978) were the intermediatic form of the subtests on whole number computation, numeration, principles of computation, sentence solving, one-step verbal problem solving, and meanings and models for each of the four basic arithmetic operations, i.e., addition, subtraction, multiplication and division. Each of the subtests was divided into groups of 3-4 items which were hypothesized as constituting a homogeneous domain. Macready and Merwin (1973) have suggested that the construction of item domains should be based on the homogenity of item content and the internal consistency of examinee responses. If the items in domain are internally consistent and test the same content, then it is reasonable to assume that mastery of all items in a domain is an all-or-none process.

The following table shows the number of domains for each subtest being researched.

SE 029 850

Subtest	Operation +	-	x	÷	Total
Computation	2	3	3	4	12
Meanings/Modles	(5 total)		(7 total)		12
Sentence Solving	[3 total]		[3 total]		6
Verbal Problem Solving	1	1	1	1	4
Principles of Computation	1	1	1	1	4

In addition, there were 7 domains which made up the subtest on numeration. There were a total of 45 disjoint domains. In the sentence solving test there were a total of 24 sentences with three factors being considered as possible determinants of homogenity. One factor was the indicated operation, another was the size of the constants in the problem and the third was the position of the placeholder in the sentence.

Table 2 gives some examples of the type of sentences used in this section.

Table 2

Sample Sentences

Operation	Basic Fact	2-digit Constants
+	9 + □ = 14	35 + □ = 50
-	□ - 9 = 8	76 - 29 = □
x	6 x 9 = □	□ x 15 = 90
÷	48 ÷ □ = 6	□ ÷ 12 = 24

As noted above 3-5 items were included in each hypothesized homogeneous domain. The test was administered to 515 fourth and fifth grade students in 3 suburban Washington, D. C. schools.

The diagnostic validity of the MDAT was defined as fit to a latent trail model which is based on an assumption of two extreme types in the population. This probahilistic model based on a compound binomial accounting for intrusion and omission rates per item is presented in detail elsewhere. (Dayton and Macready, 1976) and its application in the assessment of mastery is discussed in yet another source (Macready and Dayton, 1977).

The probabilistic model is based on dichotomously scored items (0,1) which are grouped into domains. The domain of items is constructed so that if a student possesses the skill or concept being tested by the domain, he will be able to correctly respond to all items in the domain and would be classified as a master of the domain. Conversely, if a student is a non-master of the skill or concept, then he should not be able to answer any items in the domain correctly and will have a domain score of 0.

The domains were determined a priori on the basis of the rationale underlying Wilson's (1976). Content Taxonomy and Structures. Fit to the extreme types model was then used as the criterion for diagnostic validity, i.e., are the domains homogeneous enough to identify students as either masters or non-masters for all domains being researched.

The factors that are taken into account by the probabilistic models being used in this study are the errors made by students which account for total scores other than zero (for non-masters) or four (for masters on a four item domain). In the ideal situation, a domain is homogeneous and the population is made up of two disjoint sets of subjects, masters and non-masters. The subjects who are masters are

able to do all four items in the domain correctly (producing a score of 4) while all the non-masters have a score of zero having incorrectly answered each item in the domain. In the case of 50% masters in the population, this situation would result in a symmetric bimodal distribution with no scores at 1, 2 or 3.

However, in most situations, this ideal result is not obtained. Some subjects who have acquired the skill somehow have a lapse of memory and rather than a score of 4, the score is 3 or 2. In like manner, some subject who has not acquired the concept is able to guess and rather than a score of 0, his score is 1 or 2. If one strictly adheres to a mastery score of four for the domain then, the subject who "forgot" will be misclassified as a non-master, which will be an error of classification. The model under discussion takes into account the probability that a master will "forget" a given item. This is referred to as beta, the omission or "forgetting" error associated with each item. This error rate is a conditional probability. It is the likelihood that a subject who has acquired the concept will be unable to come up with the correct answer. It is an oversimplification to call beta the "forgetting" error since many factors are contributing to the lack of a correct response.

The other misclassification error that is possible for a mastery/non-mastery dichotomy is that of the non-master being classified as a master in the case where he has "guessed" well enough to produce a mastery score. This produces a "guessing" or intrusion error, referred to as alpha, an error rate associated with each item. Alpha is also a conditional probability. It is the likelihood that a subject who has not acquired the concept will be able to produce a correct answer on a specific item in the domain.

In addition to the estimated values for alpha and beta, each model estimates the proportion of masters in the population. This parameter is theta for both models.

In summary then, there are three parameters estimated under Model I; theta - the proportion of masters; alpha - the intrusion or "guessing" error; and beta - the omission or "forgetting" error. Model II estimates 2K + 1 parameters for K item domain. These are: theta - the proportion of masters; $alpha_i$ for i = 1 to K - the intrusion error for each item; and $beta_i$ for i = 1 to K - the omission parameter for each item.

In determining a fit of the data to Model I, because only three parameters are estimated, the difficulty in finding a solution is not as great as it is for finding a solution to Model II where 2K + 1 parameters are being estimated. Thus, it is possible to have a data set such that solution to Model II will not be found. Conversely, if the items in a domain have very widely disparate values for alpha and/or beta, the restrictive assumption of Model I is not met, and a solution for Model I will not be found. For some data sets, a solution to Model I which is a mathematical artifact will occur and estimates of theta under the two models will be quite different.

Since theta is an estimated proportion of masters, and alpha and beta are probabilities, each parameter is restricted to a zero-one range. In finding a solution under the models, a parameter may exceed these limits. When this occurs, the values will be indicated as having exceeded the limits. For further discussion of this point, see Dayton and Macready (1976).

The fit of the model to the data is assessed by the standard (Pearson) chi-square goodness-of-fit test. The "expected" or predicted frequencies for each of the score patterns of zeros and ones for the domain are compared to the observed frequency for each pattern of zeros and ones. This Pearson chi-square statistic is evaluated as chi-square with degrees of freedom equal to 2^K-q-1, where q is the number of parameters estimated under the model (e.g., q = 3 in Model I).

Summary of Results for MDAT

Forty-five non-overlapping domains were investigated using the extreme types

model to determine homogeneous domains. Twelve other domains which were determined a priori were also investigated.

Results from four domains were deemed invalid because of wide discrepancies in the estimated proportion of masters as determined by the two models. Of the remaining 41 non-overlapping domain, 24 domains were homogeneous as defined by fit of the data to one or both of the models. Twenty-one of the homogeneous domains showed significantly better fit to Model II indicating that item error rates are differential across items within these domains. Twenty-one of the 24 homogeneous domains also were reliable as defined by consistency of mastery classification for comparable groups.

The table below lists domains which are homogeneous and non-homogeneous by test category:

Homogeneous Domains by Test Category

Category	Homogeneous Domains	Non-homogeneous Domains
IVD1 - 4 (Computation)	1,2,3,7,8,11,12	4,5,6,9,10
IVB1 - 2 (Meanings/Models of Addition & Subtraction)	13,15,17	14,16
IVE1 - 4 (Sentence Solving)	18C,18(BF)*,18(N-BF)* 19C,19(BF)*,19(N-BF)	18A,18B,18(+), 18(-),19A,19B 19(x),19(÷)
IVF1 - 4 (Verbal Problem Solving)	21	20,22,23
IVD1 - 4 (Principles of Computation)	27	24**,25**,26**
IVB3 - 4 (Meanings/Models of Multiplication & Division)	30,31,34	28,32,33,29**
IVA2a (Notation Systems)	35,36,38,	37,39,40,41

The reasons suggested as possible explanations for lack of homogeneity of domains varied with each test category. In general, lack of homogeneity could be accounted for by the number of regroupings required in computation, the presence of zero in a multidigit numeral, basic facts/non basic facts in sentence solving and relating the inverse operations of addition/subtraction or multiplication/division with terms, models and number sentences.

The results of the group-administered MDAT as summarized above provide some evidence for the validity of the MDAT as a diagnostic test. For those domains which were found to be homogeneous and reliable, the results do discriminate between masters and non-masters on given content objectives, providing support for the diagnostic validity of the MDAT. About half (21 of 41) of the domains were both homogeneous and reliable.

Analysis of Structures

Since the number of domains which were found homogeneous and reliable was only about half of all the domains tested, the investigation of the structure of hierarchical relationships was restricted to a few of the posited hierarchies. The results of these investigations will be discussed briefly.

Several theoretical/logical relationships between domains were to be investigated as part of this study. One structure to be investigated involved domains on computation, notational systems and principles of computation. The structure to be investigated is one which includes domains with items on (1) division computation with regrouping (2) non-standard numerals, and (3) principles of division computation. The specific domains are domain 11, division with a multiple of ten; domain 12, two digit division with regrouping; domain 38, non-standard to standard numerals; and domain 27, principles of division computation. Each of these domains was found both homogeneous and reliable. The theoretical/logical structure of these four domains was

hupothesized as being a convergent hierarchy with mastery in domains 11 and 12 on the same level; the second level being mastery of changing non-standard numerals to standard numerals and on the third level, mastery of the principles of division computation. This structure is diagrammed in figure 1.

Figure 1

Mastery scores for each of the domains were determined so that the proportion of misclassified masters was approximately equal for each domain. This gave a cut score of 3 for domains 11, 12, 38 and 27.

The table below shows the parameter estimates for the validation of the structure, assuming equal guessing and forgetting probabilities, designated alpha and beta respectively. Theta J is the proportion of the Jth true score pattern.

Table 134

Parameter Estimates for Structure 1

Parameter	Estimated Value	Standard Error
Theta 1	.554	.0269
Theta 2	.058	.0137
Theta 3	.104	.0166
Theta 4	.205	.0208
Theta 5	.061	.0127
Alpha	.029	.0065
Beta	.001	.0144

chi-square = 16.30 with 8 degrees of freedom, p = .041.

In the chi-square summary table below, the patterns do not represent item scores, but rather mastery of a domain. Thus, the first digit in the pattern represents a master/non-master of domain 11 (2-digit division computation with a multiple of ten), the second digit represents a master/non-master of domain 12 (2-digit division with regrouping), the third digit represents domain 38-non-standard numerals and the fourth digit represents the mastery/non-mastery on domain 27-principles of division computation.

The chi-square summary table below shows how closely the empirical results match the true score patterns determined from the theoretical/logical structure. These results support the contention that students who are able to do division computation involving a two-digit divisor with regrouping of the dividend are not necessarily students who understand the application of place value concepts to the computational process of division with regrouping.

Table 135

Chi-square Summary Table for Structure 1

Pattern	Observed N	Predicted N	Chi-square
0000	* 218	212.3303	.1514
1000	* 29	29.2846	.0028
0100	* 45	47.2061	.1031
1100	* 82	85.4330	.1379
0010	7	6.2549	.0888
1010	0	.8627	.8627
0110	3	1.3905	1.8625
1110	* 28	27.9734	.0000
0001	0	6.2549	6.2549
1001	2	.8627	1.4994
0101	2	1.3906	.2670
1101	6	2.5167	4.8211
0011	0	.1843	.1843
1011	0	.0254	.0254
0111	0	.0410	.0410
1111	* 8	7.9889	.0000

* = True Score Pattern

These results support the contention that for the operation of division with regrouping, the empirical results are in accord with the theoretical/logical structure of relationships between computation, numeration concepts and principles of computation as diagrammed above.

Implications for Future Research - Extreme Types Model

The results of the application of the Dayton and Macready extreme types of model to the MDAT revealed that the model is a very useful tool in the development of a domain-referenced test. The intrusion and omission parameters which are estimated for each item in the domain under Model II provide valuable information about possible faulty items as well as providing clues about how to better structure a domain of items. While the statistics resulting from the application of the models are valuable in determining homogeneous domains, this information must be taken in context. For some domains the data fit Model I or Model II but the results were invalid because of widely discrepant estimated values for the proportion of masters in the population under the two models. Thus, the "fit" of the data to a model was not accepted as the sole criterion of a homogeneous domain of items.

Another point to be considered in making up a homogeneous domain is the content covered by each item in the domain. As was noted above, the data which are the input for the models are only score frequencies of pattern frequencies. Therefore, it is conceivable that four unrelated items could be found such that the data would fit one or both of the models implying that a homogeneous domain exists for these unrelated items. Content specialists should also be used to determine what are reasonable limits for the intrusion and omission parameters for specific items within domains. These limits for the item error rates could then help in reaching decisions on whether a domain should be restructured or subdivided. The clinical experience of content specialists in finding possible explanations for the lack of homogeneity of

a domain of items is invaluable. The statistics from the model can provide clues to student test behaviors which may contradict what is generally believed about student learning. However, the statistics alone are not enough for suggesting possible answers to the how and why of the observed patterns of domain scores. The models also provide an excellent way to examine the same items in several ways to see if hypothesized relationships and/or student strategies do hold up for different operations and/or sets of numbers.

References

Hambleton, R.K. and Eignor, D.R. Guidelines for evaluating criterion-referenced tests and test manuals. Journal of Educational Measurement, 1978, 15, 321-327.

Macready, G.B. and Dayton, C.M. The use of probabilistic models in the assessment of mastery. Journal of Educational Statistics, 1977, 2, 99-120.

Popham, W.J. Criterion-referenced measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Sadowski, B.R. An investigation of the validity, accuracy and consistency of prescriptions derived from a domain-referenced diagnostic arithmetic test and interview protocols. Unpublished doctoral dissertation, University of Maryland, College Park, 1978.

Domain 11 - Division With a Multiple of Ten

The items in domain 11 are:

1) $20\overline{)4080}$ 2) $50\overline{)2550}$ 3) $30\overline{)6930}$ 4) $40\overline{)1680}$

Table 34

Model I Results for Domain 11

Parameter	Estimated Value	Standard Error
Theta	.492	.0257
Alpha	.075	.0115
Beta	.209	.0164

Chi-square = 6.71 with 1 degree of freedom, $p = .009$.

Table 35

Model II Results for Domain 11

Parameter	Estimated Value	Standard Error
Theta	.505	.0251
Alpha 1	.034	.0137
Alpha 2	.105	.0226
Alpha 3	.114	.0235
Alpha 4	.051	.0182
Beta 1	.389	.0330
Beta 2	.171	.0265
Beta 3	.139	.0247
Beta 4	.118	.0244

Chi-square = 16.23 with 6 degrees of freedom, $p = .013$.

Domain 12 - Two-digit Division With Regrouping

The items for domain 12 are.

1) $13\overline{)433}$ 2) $21\overline{)883}$ 3) $18\overline{)756}$ 4) $24\overline{)846}$

Table 36

Model I Results for Domain 12

Parameter	Estimated Value	Standard Error
Theta	.437	.0240
Alpha	.039	.0733
Beta	.204	.0165

Chi-square = 3.86 with 1 degree of freedom, p = .045.

Table 37

Model II Results for Domain 12

Parameter	Estimated Value	Standard Error
Theta	.436	.0237
Alpha 1	.021	.0106
Alpha 2	.079	.0176
Alpha 3	.039	.0139
Alpha 4	.019	.0094
Beta 1	.178	.0286
Beta 2	.172	.0277
Beta 3	.125	.0250
Beta 4	.337	.0340

Chi-square = 10.98 with 6 degrees of freedom, p = .221

Domain 27 - Principles of Division

The items of this domain are:

IN THE EXAMPLE $25\overline{)5746}$, THE QUOTIENT (OR ANSWER) WILL BE:

A ☐ BETWEEN 2 AND 3
B ☐ BETWEEN 20 AND 30
C ☐ BETWEEN 200 AND 300
D ☐ BETWEEN 2000 AND 3000

WHEN DIVIDING $42\overline{)5642}$, IN WHICH PLACE WILL THE FIRST FIGURE OF THE ANSWER BE WRITTEN?

A ☐ ONES B ☐ TENS C ☐ HUNDREDS D ☐ THOUSANDS

THE GREATEST REMAINDER YOU SHOULD HAVE FOR THIS EXAMPLE $9\overline{)\square\square\square}$ IS:

A ☐ 1 B ☐ 8 C ☐ 9 D ☐ CAN'T TELL

IN THE EXAMPLE BELOW, THE NUMERAL TO WHICH THE ARROW POINTS MEANS:

$$\begin{array}{r} 5 \\ 77\overline{)40674} \\ -385 \\ \hline 21 \leftarrow \end{array}$$

A ☐ 21 ONES
B ☐ 21 TENS
C ☐ 21 HUNDREDS
D ☐ 21 THOUSANDS

Table 95

Model I Results for Domain 27

Parameter	Estimated Value	Standard Error
Theta	.516	.1407
Alpha	.185	.0489
Beta	.477	.0528

Chi-square = 2.12 with 1 degree of freedom, p = .146.

Table 96

Model II Results for Domain 27

Parameter	Estimated Value	Standard Error
Theta	.486	.0964
Alpha 1	.284	.0574
Alpha 2	.210	.0494
Alpha 3	.270	.0431
Alpha 4	.151	.0267
Beta 1	.193	.0881
Beta 2	.292	.0908
Beta 3	.351	.0720
Beta 4	.707	.0488

Chi-square = 8.36 with 6 degrees of freedom, p = .213.

Domain 38 - Non-standard - Standard Num erals

The items included in domain 38 are:

WRITE THE CORRECT NUMERAL IN EACH BLANK.

A) 7 TENS + 15 ONES = _____

B) 42 TENS AND 23 ONES = _____

C) 965 = 15 ONES + ____ TENS + 9 HUNDREDS

D) 2163 = ____ HUNDREDS + 15 TENS + 3 ONES

Table 122

Model I Results for Domain 38

Parameter	Estimated Value	Standard Error
Theta	.293	.0436
Alpha	.080	.0119
Beta	.495	.0382

Chi-square = 2.41 with 1 degree of freedom, p = .121.

Table 112

Model II Results for Domain 38

Parameter	Estimated Value	Standard Error
Theta	.273	.0274
Alpha 1	.084	.0270
Alpha 2	.006	.0048
Alpha 3	.081	.0155
Alpha 4	.011	.0060
Beta 1	.003	.0126
Beta 2	.513	.0539
Beta 3	.376	.0544
Beta 4	.779	.0389

Chi-square = 18.88 with 6 degrees of freedom, p = .001.